# Numerical Weather Prediction

## 4D-Var and the Butterfly Effect.
## Statistical 4D-Var for a wide range of scales.

Forecasting Research Technical Report No. 481

Andrew C. Lorenc

*email: nwp_publications@metoffice.gov.uk*

©Crown Copyright

# 4D-Var and the Butterfly Effect.
# Statistical 4D-Var for a wide range of scales.

*Andrew C. Lorenc 2006-05-15*

# ABSTRACT

I review the limits of 4D-Var, particularly its ability to work well for a wide range of scales, and discuss the relationship to the atmospheric butterfly effect. As a concrete example, I consider how 4D-Var might be applied to a global, convective-scale model. Deterministic 4D-Var, finding the most likely model evolution to fit observations, does not work in the limit of high resolution. Statistical 4D-Var, minimising the mean-square errors and estimating the mean of possible states, might work. It requires a means of forecasting the evolution of the best estimate state; the parametrisation of uncertain scales in such a model is discussed. Given the difficulties of building such a model, a seamless NWP system, based on 4D-Var and capable of forecasting from convective to global scales, is unlikely. A more feasible way of getting a best estimate forecast is by averaging deterministic forecasts. This could be combined with an incremental 4D-Var treating only the well observed, synoptic scales. Such a system could give good forecasts of forced convection, and of the average convection. In areas with sufficient convective-scale observations, we could use limited-area NWP systems, nested in the global system and not attempting to improve on the synoptic scales. These might use 4D-Var designed for the convective-scale.

ABSTRACT ..... 1
1. Introduction ..... 1
2. Deterministic 4D-Var: estimation of the most likely state ..... 2
3. Statistical 4D-Var: the "best estimate" state ..... 3
4. Allowing for model error ..... 4
5. Forecasting the best estimate mean state ..... 4
6. Forecasting the best estimate of a finite perturbation ..... 5
7. Discussion ..... 6
References ..... 7
Figures ..... 9

## *1. Introduction*

Computer power is increasing at such a rate that it will soon be routinely possible to run global numerical weather prediction (NWP) forecasts that resolve convective-scale systems. The explicit representation of such scales, even though they are not fully observed, should improve forecasts of larger scales, through the better representation of the effects of deep convection.

Four-dimensional variational data assimilation (4D-Var) has been very effective in operational global NWP (Rabier 2006). 4D-Var also works well for convective scales (Sun 2006). This paper's goal is to consider the theoretical limits of the 4D-Var technique, rather than to propose a particular design for a particular application. But, since assimilation for a wide range of scales is a key issue, we use as an example the possibility of applying 4D-Var to a global convective-scale NWP system.

To approach the optimal solution, 4D-Var should consider observations in a time-window of similar length to the life-cycle of the scales being considered – Fisher *et al.* (2006) suggest that for global NWP about 5 days are needed. However convective scales are not predictable for this long; they grow rapidly until limited by nonlinear processes not well handled by standard 4D-Var methods. This paper considers how 4D-Var can be modified to handle a flow which possesses many scales of motion. This is closely related to the problem of atmospheric predictability (Lorenz 1969, Tribbia and Baumhefner 2004) which has become

known as the atmospheric butterfly effect[1] (e.g. Palmer 2005) and is referred to in the title. The applicability of conventional deterministic 4D-Var was discussed by Tanguay *et al.* (1995); they showed that for a given assimilation period there is a scale below which useful initial conditions cannot be obtained. This is reviewed in section 2.

In sections 3 to 6 we discuss the statistical 4D-Var approach of Lorenc (2003) to see how it might be made convergent in the sense that it continues working, converging to the correct result for well-resolved scales, as the resolution of the model used is increased. As well as coping with a model representing a wide range of scales, an ideal 4D-Var should cope with a wide range of observation densities. We seek a method which does at least as well, for synoptic scales, assimilating a synoptic-scale observation network into a convective-scale model as into a synoptic-scale model, while at the same time being capable of assimilating convective-scale observations if they exist. It is not clear that convergence is possible for NWP models of the atmosphere (as opposed to idealised problems) for the scales we are interested in, because significant new features become resolved as the resolution increases. The topic is of current interest, in discussions of seamless NWP and climate models (WCRP 2005). Here, we try to start to extend that discussion to 4D-Var.

Finally, section 7 concludes that a seamless 4D-Var meeting the above goals is unlikely to be built. A more practical vision is presented instead.

## 2. Deterministic 4D-Var: estimation of the most likely state

In its normal Bayesian derivation (Purser 1984, Lorenc 1986), an expression for the probability distribution function (PDF) is:

$$P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) \propto P\left(\mathbf{x}\right) P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right) \tag{1}$$

Notation follows Ide *et al.* (1997) with an extension that, to avoid explicit summations over time, underlining represents a time-sequence of the variable. The prior PDF for model state $\mathbf{x}$ is a Gaussian with mean $\mathbf{x}^b$ and covariance $\mathbf{B}$:

$$P\left(\mathbf{x}\right) \propto \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right)\right) \tag{2}$$

The observations in the time-window, $\underline{\mathbf{y}}^o$, are usually assumed to have Gaussian errors in observation-space, with covariance $\underline{\mathbf{R}}$, uncorrelated with background errors:

$$P\left(\underline{\mathbf{y}}^o\middle|\mathbf{x}\right) = P\left(\underline{\mathbf{y}}^o\middle|\underline{\mathbf{y}}\right) \propto \exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right) \tag{3}$$

$\underline{\mathbf{y}}$ represents the estimate of the observations calculated from $\mathbf{x}$:

$$\underline{\mathbf{y}} = \underline{H}\left(\underline{M}\left(\mathbf{x}\right)\right) \tag{4}$$

$\underline{M}$ is a forecast from $\mathbf{x}$ (which is valid at the start of the time-window) to the observation times; $\underline{H}$ is the observation operator which calculates from this forecast the predicted values of the observed variables. Errors in (4) are ignored, except that errors of representativity are included in $\underline{\mathbf{R}}$ (Lorenc 1986). $\underline{M}$ and $\underline{H}$ are nonlinear, so the PDF obtained by substituting (2), (3) and (4) into (1) is not Gaussian. With very little justification other than convenience of manipulation, it is usually assumed that the desired estimate is the $\mathbf{x}$ which maximises $P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right)$, or equivalently minimises

$$J\left(\mathbf{x}\right) = \tfrac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right) \tag{5}$$

subject to (4). This is commonly called the 4D-Var penalty function. It can be elaborated, for instance to allow for non-Gaussian observation errors or imperfect forecast model. For practical NWP models the process of finding the minimum is always approximated. The gradient of J with respect to $\mathbf{x}$ is given by

$$\nabla_{\mathbf{x}} J\left(\mathbf{x}\right) = \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \underline{\mathbf{M}}^{*}\underline{\mathbf{H}}^{*}\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right) \tag{6}$$

[1] The butterfly effect describes a finite limit to predictability due to the existence of increasingly small scales. It is different from the chaotic limit to predictability exhibited for instance by the three parameter Lorenz (1963) model, studied in a data assimilation context in papers such as Pires *et al.* (1996).

where $\underline{\mathbf{M}}^*$ and $\underline{\mathbf{H}}^*$ are the adjoints of $\underline{M}$ and $\underline{H}$. (5) and (6) are used in an iterative descent algorithm. Practical problems in finding the minimum, due to the nonlinear and sometimes non-differentiable $\underline{M}$ and $\underline{H}$, and problems of multiple minima, have been discussed in the literature. But there has been surprisingly little discussion of the fact that the minimum of this penalty function is not the best estimate anyway. The statistical 4D-Var approach discussed in the next section addresses all these issues.

Deterministic 4D-Var is in principle finding the most likely trajectory from a physically based model, given the observations. Real atmospheres exhibit intermittent threshold processes such as clouds, so it is natural that NWP models should represent these by tests and switches. These cause problems when defining tangent-linear and adjoint models (e.g. Xu 1996).

The problems of multiple minima are a fundamental to the extension of deterministic 4D-Var to many scales of motion (Tanguay *et al.* (1995) figure 12). To see why, consider figure 2 of Lorenz (1969), or the recently updated figure 1 of Tribbia and Baumhefner (2004). This shows that errors, or in our case small perturbations, grow with timescales which decrease with their horizontal scale. So for a give 4D-Var time-window, sufficiently small horizontal scales will be uncertain. As many atmospheric motions are wavelike, this means that features a cycle or more incorrect will give spurious fits to particular observations. This will give spurious minima in the penalty function. This effect was illustrated, for one particular time-scale, by Pires *et al.* (1996) figure 4. As we include smaller scales in a 4D-Var whose time-window is fixed by the largest waves, eventually this effect will occur.

## 3. Statistical 4D-Var: the "best estimate" state

The problem discussed in the previous section can be avoided if we seek a more appropriate "best" estimate. To define what is best, it is necessary to specify a cost function measuring the additional costs incurred by issuing an imperfect forecast. The simplest is a quadratic cost: the expected root mean square error (RMSE) is minimised by the mean of the PDF[2]:

$$\mathbf{x}^a = \int \mathbf{x} P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) d\mathbf{x} \Big/ \int P\left(\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) d\mathbf{x} \qquad (7)$$

This is a function of the whole PDF, rather than just the derivatives near the minimum used in deterministic 4D-Var. It is not amenable to calculation for PDFs as complicated as (1) for a full NWP model. However, rather than minimising the wrong equation (5), it may be better to find an approximate solution to (7). Lorenc (2003) suggested one way to do this, based on the fact that the mode of a PDF is also its mean, as long as the distribution is Gaussian. It is the nonlinear evolution of the prior PDF $P(\mathbf{x})$ by $\underline{M}$ which makes (1) non-Gaussian (Fig. 1). So let us instead approximate that evolution by $\overline{\underline{M}}$, which predict the evolution of the mean, and $\tilde{\underline{\mathbf{M}}}$ which gives linear best estimates of the evolution of perturbations about this mean (Fig. 2). If necessary, a similar strategy could be applied to $\underline{H}$, $\overline{\underline{H}}$ and $\tilde{\underline{\mathbf{H}}}$. The whole PDF (1) (rather than its behaviour near its mode) is then approximated using an incremental approach about a guess $\mathbf{x}^g$ of the ensemble mean (Courtier *et al.* 1994):

$$\begin{aligned} P\left(\delta\mathbf{x}\middle|\underline{\mathbf{y}}^o\right) \propto & \exp\left(-\tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T \mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)\right) \\ & \exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right) \end{aligned} \qquad (8)$$

$$\underline{\mathbf{y}} = \tilde{\underline{\mathbf{H}}}\tilde{\underline{\mathbf{M}}}\delta\mathbf{x} + \overline{\underline{H}}\left(\overline{\underline{M}}\left(\mathbf{x}^g\right)\right) \qquad (9)$$

Since (9) is linear in $\delta\mathbf{x}$, (8) is Gaussian. So if we use the same descent algorithm approach to efficiently approximate the minimum of

$$\begin{aligned} J\left(\delta\mathbf{x}\right) = & \tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T \mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right) \\ & +\tfrac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T \underline{\mathbf{R}}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right) \end{aligned} \qquad (10)$$

[2] Lorenc (2002), discussing the definition of "best" in the context of non-Gaussian PDF, suggested a Gaussian shaped cost function. This is optimised by the mean for a very broad Gaussian, and the mode for a very narrow delta function. In between it is possible to adjust it to pick the largest peak of a bi-modal PDF.

subject to (9), we are also approximating the mean of (8). By construction (10) is quadratic, so we avoid the minimisation problems which plague (5).

## 4. Allowing for model error

Thus far, for simplicity of presentation, I have not included terms allowing for forecast model error. However these are essential for long-window 4D-Var (Fisher *et al.* 2006), and therefore essential for the multi-scale 4D-Var we are considering, since any time-window is long for short enough space scales. Lorenc (2003) discussed how model error can be regarded as just another component of a four-dimensional covariance model, and included in the control variable transforms used in practical 4D-Var schemes. Assuming the model and background errors are independent, we approximate the model error distribution by a Gaussian with mean $\underline{\boldsymbol{\eta}}^g$ and covariance $\underline{\mathbf{Q}}$, and add an additional model-error control variable $\delta\underline{\boldsymbol{\eta}}$.

$$
\begin{aligned}
P\left(\delta\mathbf{x},\delta\underline{\boldsymbol{\eta}}\middle|\underline{\mathbf{y}}^o\right) \propto \exp&\left(-\tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T\mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)\right)\\
&\exp\left(-\tfrac{1}{2}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)^T\underline{\mathbf{Q}}^{-1}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)\right) \qquad (11)\\
&\exp\left(-\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)\right)
\end{aligned}
$$

The definition of $\tilde{\mathbf{M}}$ and $\bar{M}$ are extended to include addition of model error terms:

$$
\underline{\mathbf{y}}=\underline{\tilde{\mathbf{H}}}\tilde{\mathbf{M}}\left(\delta\mathbf{x},\delta\underline{\boldsymbol{\eta}}\right)+\underline{\bar{H}}\left(\bar{M}\left(\mathbf{x}^g,\underline{\boldsymbol{\eta}}^g\right)\right) \qquad (12)
$$

The mean of (11) is identical to its mode, so can be found by minimising a penalty function of $\delta\mathbf{x}$ and $\delta\underline{\boldsymbol{\eta}}$

$$
\begin{aligned}
J\left(\delta\mathbf{x},\delta\underline{\boldsymbol{\eta}}\right)=&\tfrac{1}{2}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)^T\mathbf{B}^{-1}\left(\delta\mathbf{x}-\left(\mathbf{x}^b-\mathbf{x}^g\right)\right)\\
&+\tfrac{1}{2}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right)^T\underline{\mathbf{Q}}^{-1}\left(\delta\underline{\boldsymbol{\eta}}+\underline{\boldsymbol{\eta}}^g\right) \qquad (13)\\
&+\tfrac{1}{2}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o\right)
\end{aligned}
$$

subject to (12).

Note that $\bar{M}$, $\underline{\bar{H}}$, $\tilde{\mathbf{M}}$ and $\underline{\tilde{\mathbf{H}}}$ should be designed so that substituting (12) into (11) gives a good Gaussian approximation to the posterior PDF which would be obtained with the accurate nonlinear models and the Fokker-Planck equation (Cohn 1993). Accurate behaviour is less important for improbable cases where the PDF is small. The initial $\mathbf{x}^g$ and $\underline{\boldsymbol{\eta}}^g$ are usually $\mathbf{x}^b$ and zero; a better Gaussian approximation to the PDF can be obtained by updating these in an outer-loop (Courtier *et al.* 1994). See Lorenc (2003) for details of the outer-loop, transformations and simplifications which can be used in a practical NWP 4D-Var system.

## 5. Forecasting the best estimate mean state

NWP systems have always filtered unpredictable scales. This is often thought of as an approximation forced by limited computers, but in some cases it has been recognised as improving forecast verification scores such as RMSE [3]. The desirability of doing this for data assimilation was recognised by Cohn (1993), who said *"If Kalman-like filters are to be used at all ... it will probably be better to use second-moment (or higher) closure filters, in which an ensemble mean equation is used"* and *"In the context of variational assimilation, these arguments may also eventually come into play ... it may well be necessary to find a way to account for the difference between current model dynamics and ensemble mean dynamics."* Cohn showed that the mean forecast equation has an extra term proportional to the local variance of the PDF. This can be thought of as an extra parametrisation for the effect of uncertain scales of motion; it can be contrasted with normal NWP model parametrisations of unresolved scales of motion.

In practice the model resolution for particular application has usually been chosen to only represent the scales that we hope to forecast and that are well observed, so there has been little difference between uncertain and unresolved scales. The uncertainty increases however in longer forecasts; this was recognised

[3] The issue whether RMSE scores properly measure forecast usefulness is touched on briefly in the discussion.

by Jakimow *et al.* (1992), who increased the diffusion as the forecast progressed, decreasing RMSE after 6 days (WGNE 1996). Selection is not only by scale: unbalance gravity waves are not predictable, so it has long been the practice to filter them by an initialisation scheme.
A simple conceptual model of the parametrisation of uncertainties is behind the above descriptions: waves whose phase is uncertain will cancel in the mean, so they should be removed from the mean forecast. Features whose position is uncertain should be filtered on a scale determined by the uncertainty. This filtering can be gradual, with a timescale related to the predictability of the features. Then the effect of the filtered features on the larger scales needs to parametrized.
Note that a filtered NWP model would not give an accurate prediction for the filtered scales, even if the input field happened to be accurate. So the filtering is an additional contribution to the model error terms discussed above.
The mean forecast behaviour is not determined by the mean state alone – the amount of filtering is determined by the uncertainty. Unless the uncertainty can be assumed a fixed function of scale, or a function of time as in Jakimow *et al.* (1992), this requires a prediction of the PDF's variance. 4D-Var does not naturally provide this. It may be necessary to use additional techniques for estimating the variance, for instance from an ensemble forecasting system, or from simplified methods for predicting the variance (e.g. Cohn 1993).

It is more difficult to apply these ideas to an NWP system which deliberately represents significant features whose details are only sometimes uncertain, such as a global convective scale model. In some areas, near a weather radar for instance, the convection might be well observed, so the variance of the PDF and hence the extra filtering terms will be small. Convective cells should be retained in the forecast for as long as they remain predictable. In other areas details of the positions of convective cells are lacking. Here they should be removed by filtering and their effect replaced by a large-scale convective parametrisation. This hybrid state will appear unphysical, but it is the best forecast by RMSE measures, and therefore in principle should be used in a 4D-Var designed to make the best estimate.

4D-Var for a wide range of scales was studied in a more extreme context by Köhl and Willebrand (2002), who wanted to fit long-term climate observations to an eddy-resolving ocean model. They came to the same conclusion as here: the need to use a model that describes PDFs or mean values independently of special realisations. However they decided that such models were not available, so instead they post-processed a full model run to produce such a mean trajectory. They used time-averaging; space-averaging or an ensemble average could be used as well.

## 6. Forecasting the best estimate of a finite perturbation

Lorenc (1997) described plans for used a 4D-Var system based on (10), using a linear Perturbation Forecast (PF) model to approximate the evolution of finite-sized perturbations, rather than the tangent-linear (TL) model which predicts the evolution of infinitesimal perturbations. It has proved useful, when diagnosing the performance of 4D-Var schemes, to test the PF model against the full nonlinear model for the evolution of typical finite perturbations (Ballard *et al.* 2006, Radnóti *et al.* (2005)).

The PF model is only used to estimate errors and weight information; it does not propagate information (Lorenc 2003). So it can be run at lower resolution, saving computer costs (Courtier *et al.* 1994). This strategy works when the scales omitted from the PF model are not determined directly by the observations (Köhl and Willebrand 2002).

In considering the best estimate for a finite perturbation over a long window, similar considerations of predictability need to be considered. Tanguay *et al.* (1995) showed that small increments grow exponentially, most rapidly in regions of strong gradients. In linear models this exponential growth continues indefinitely; in reality the growing modes will disrupt the gradients and reduce the growth. Mean perturbation growth ceases when the amplitude squared reaches twice the climatological error variance. To approximate this effect, the growth rates from the PF model need to be reduced for finite perturbations. The need for such damping has been seen in the extended Kalman filter (Miller *et al.* 1993), and in long window ocean 4D-Var (Köhl and Willebrand 2002, Hoteit *et al.* 2006). As in the equation for the mean forecast, the additional term will depend on the variance of the PDF - when all likely perturbations are small the term is not needed.

Rather than modifying the adjoint as discussed above, Lea *et al.* (2000) tried to avoid the exponential growth of some modes by a process of ensemble averaging of short trajectories. It is not clear how to do

the averaging in a general case, but the approach may be appropriate when the climate sensitivities being calculated are desired in their own right. For 4D-Var a simplified PF and adjoint model is likely to be more cost-effective (Lorenc 1997, Köhl and Willebrand 2002).

This concept of designing the PF model to give the best linear estimate for a range of finite perturbations is particularly helpful for parametrisations. For example Sun and Crook (1997) showed that physical equation for evaporation of rain has very steep gradients for low rates, but these only apply for a small range of model values. Payne (2005, personal communication) showed how such high gradients, and the discontinuities discussed earlier (Xu 1996), are smoothed and regularised by a PF model optimised for a distribution of perturbations. In this context, it is helpful to consider the spread of the regularising distribution of perturbations. In the initial application of (13), $\mathbf{x}^g$ is usually set to $\mathbf{x}^b$, and the PF model should give a good approximation for the whole background PDF. Sometimes the background will have quite large errors, passing thresholds such as the onset of cloud and rain. So the PF model should give average behaviour across the threshold; it will represent gradients reflecting the closeness to cloud and rain even when the deterministic prediction from $\mathbf{x}^b$ has none. This should facilitate the assimilation of observations of cloud and rain[4]. In later iterations of the outer loop, $\mathbf{x}^g$ should be closer to the best estimate. So the PF model need only give a good approximation over the range of the posterior analysis PDF. If the algorithm has been successful, in the regions with observations the analysis error will be considerably less than the background error, so averaging across thresholds should be less important.

## 7. Discussion

Deterministic 4D-Var is unlikely to work well simultaneously for a wide range of scales. Statistical 4D-Var, using a mean forecast equation, might be able to work because it filters uncertain scales. The need for such filtering has not been very apparent in practical NWP 4D-Var systems because they tend to be designed for and applied to specific applications and scales which are reasonably well observed. 4D-Var has been very effective in global NWP data assimilation; here the scales represented have been limited by available computer power. 4D-Var also works for convective scales; here implementation has concentrated on regions where the scales of interest are relatively well observed by Doppler radar. Fisher *et al.* (2006) showed that a model error term acts as a "forgetting factor" enabling a long-window 4D-Var to approximate a Kalman filter, without employing a mean forecast equation. But their tests were with an idealised model with over half the points observed every 6 hours[5].

This paper is based on the assumption that a useful forecast can be defined by the mean of the PDF (the best estimate) together with an estimate of the variance of the PDF. We also assume that using a high resolution is worthwhile. It is then worth investing much of the available computer power in a good forecast of the PDF mean. (This is not literally a deterministic forecast, since errors are admitted and allowed for both in the forecast equations and by implicitly "dressing" the mean forecast with a PDF of possible errors.) The same assumptions then lead to statistical 4D-Var, using the mean forecast model, as a good way of assimilating observations. Note that 4D-Var gain its efficiency, and hence its improvement over other assimilation methods, by never explicitly representing error covariances (Lorenc 2003). As for the dressed forecast, it is worth investing computer resources in 4D-Var when a good estimate of the PDF mean is of high value. These assumptions only apply if the PDF is well approximated by a Gaussian - this is reasonable for the short-period forecasts at relatively well observed scales typical of current NWP assimilation systems. In other situations, where RMSE is not a measure of a useful forecast, an ensemble-based assimilation and forecast system might be more appropriate. (Practical ensemble DA systems based on the Kalman filter currently make Gaussian assumptions, and are unable to deal with a wide range of scales. It is not at all clear that ensemble DA would be any better than 4D-Var for global convective-scale NWP.)

[4] Deterministic 4D-Var has difficulties fitting such observations, since the gradient of zero cloud or rain is also zero, so the descent algorithm does not know which way to go to improve the fit to non-zero observations.

[5] The need for the damping terms in the mean forecast model can be seen in their model if only a single observation at the end of the time window is provided. Because of the continuous growth in their linear model (with Lyapunov doubling time of 2.1 days), extending the 4D-Var window will provide additional degrees of freedom able to fit the observation with smaller and smaller increments. There is no convergence to a sensible result; it instead is seeking smaller and smaller butterflies, earlier and earlier. With damping such as to stop growth when the PDF spread is large, this would not happen.

The ideal, global, convection-resolving 4D-Var system, part of a seamless prediction system from convective to global scales, whose consideration prompted this work, might be built as follows. Build a long-window 4D-Var scheme (following Fisher *et al.* 2006) so as to get the best assimilation possible of global and synoptic scales. Build a separate system to estimate the uncertainty variance in different scales, as a function of time and place. Add selective damping, proportional to the uncertainty variance, to the forecast model used in 4D-Var, so that it approximates the "best estimate" mean forecast. The model error control variables used in the long window 4D-Var should represent the scales being removed by the damping. This combination of damping and model error control variables applied during the 4D-Var window mean that the effective time-window for such scales is smaller; a short window is necessary for 4D-Var at convective scales.
Of course there would be many difficulties in building such a scheme. Not least is the problem of building a controllable, selective filter, removing uncertain scales, plus a controllable, smooth parametrisation of the mean effect of the scales which have been filtered. The difficulty of building well-behave convective parametrisation schemes is one of the justifications for wanting a global convective-scale model in the first place. NWP models are usually designed, like climate models, to represent physical processes even if they are unpredictable. This approach is also helpful in an ensemble forecast system. Each member is forecast using the physically based equations, making it much easier to base forecast model development of detailed physical studies of each process represented. Statistical 4D-Var cannot handle a model which represents significant uncertain scales. It requires a forecast model for the best estimate mean state. This is much more difficult to build. In areas without convective-scale observations it would produce a state with these scales filtered, obviating the reason for using a convective-scale model. So the vision of a seamless NWP system from convective to global scales is unlikely to be realised using 4D-Var.

A more costly, but less scientifically difficult way of forecasting the evolution of the mean of the PDF is to make an ensemble of forecasts. If it is accepted that small scales are uncertain, then a good estimate of the PDF mean can be obtained by post-processing a small ensemble with a space- and time-filter as well as an ensemble mean. Around this post-processed trajectory we could use a synoptic-scale incremental 4D-Var, with a PF model including a smooth parametrisation of convective scales. The increment could be added to each ensemble member and the assimilation cycle continued[6]. Such a system could be convergent to the best synoptic scale assimilation: increasing the resolution (and ensemble size) of the main forecast should improve the post-processed estimate of synoptic scales without altering its character; the PF model would retain its smooth parametrisation of the mean effect of convective scales at higher resolution. Where convective-scale motions are forced by synoptic scales and known topography, this system would represent them correctly. Elsewhere it should indicate the possible convection, but it would not use convective-scale observations to initialise unforced convective-scales. To do this, we could use limited-area NWP systems designed specifically for the convective-scale, nested in the global system and not attempting to improve on the synoptic scales specified by these. They could therefore use 4D-Var with a short time-window.

# References


Ballard *et al.* 2006: "A Perturbation Forecast model and its adjoint". *In preparation*.

Cohn, Stephen E. 1993: "Dynamics of short-term univariate forecast error covariances." *Mon. Wea. Rev.*, **121**, 3123-3149.

Courtier, P., J-N. Thepaut, and A. Hollingsworth, 1994: "A strategy for operational implementation of 4D-Var, using an incremental approach". *Quart. J. Roy. Met. Soc.*, **120**, 1367-1387.

Fisher, M., M. Leutbecher and G. A. Kelly 2006: "On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation." To appear in *Quart. J. Roy. Met. Soc.*, *Special Issue of papers from the 4th WMO Symposium on Data Assimilation.*

http://www.ecmwf.int/publications/library/ecpublications/_PDF/tm/401-500/tm447_rev.PDF

Hoteit, I., B. Cornuella, A. Köhl and D. Stammer, 2006: "Treating Strong Adjoint Sensitivities in Tropical Eddy-Permitting Variational Data Assimilation", *Quart. J. Roy. Met. Soc. Special Issue devoted to papers from the 4th WMO Symposium on Data Assimilation.*

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C. 1997: "Unified notation for data assimilation: Operational, Sequential and Variational" *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 181—189.


[6] The spread of the ensemble would have to be maintained independently of the 4D-Var system, for instance by stochastic physical parametrisations.

Jakimow, G., E. Yakimiw, and A. Robert, 1992: An Implicit Formulation for Horizontal Diffusion in Gridpoint Models *Mon. Wea. Rev.*, **120**, 124-130.

Köhl A. and J. Willebrand 2002: "An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models" *Tellus*, **54A**, 406-425.

Lea, D.J., M. R. Allen and T. W. N. Haine, 2000: "Sensitivity analysis of the climate of a chaotic system." *Tellus*, **52A**, 523-532.

Lorenc, A.C. 1986: "Analysis methods for numerical weather prediction." *Quart. J. Roy. Met. Soc.,* **112**, 1177-1194.

Lorenc, Andrew C., 1997: "Development of an Operational Variational Assimilation Scheme." *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 339-346.

Lorenc, A. C., 2002: Atmospheric Data Assimilation and Quality Control. *Ocean Forecasting*, eds Pinardi & Woods. ISBN 3-540-67964-2. 73-96

Lorenc, A. C., 2003: Modelling of error covariances by four-dimensional variational data assimilation. *Quart. J. Roy. Met. Soc.*, **129**, 3167-3182.

Lorenz, E. N., 1963: "Deterministic nonperiodic flow". *J. Atmos. Sci.*, **20**, 130-141.

Lorenz, E. N., 1969: "The predictability of a flow that possesses many scales of motion". *Tellus*, **21**, 289-307.

Miller, R. N., M. Ghil, F. Gauthiez, 1994: "Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems" *J. Atmos. Sci.*, **51**, 1037-1056

Palmer, T. N. 2005: "Quantum Reality, Complex Numbers, and the Meteorological Butterfly Effect" *Bulletin of the American Meteorological Society*, **86**, 519-530

Pires, C., R. Vautard and O. Talagrand, 1996: "On extending the limits of variational assimilation in nonlinear chaotic systems." *Tellus*, **48A**, 96-121.

Purser, R.J. 1984: "A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis". Preprints, 10th conference on weather forecasting and analysis. Am Met Soc. 102-105.

Rabier, Florence 2006: "Overview of global data assimilation developments in Numerical Weather Prediction centres" *Quart. J. Roy. Met. Soc. Special Issue devoted to papers from the 4th WMO Symposium on Data Assimilation.*

Radnóti, G., Y. Trémolet, E. Andersson, L. Isaksen, E. Hólm and M. Janisková, 2005: "Diagnostics of linear and incremental approximations in 4D-Var revisited for higher resolution analysis" ECMWF Tech. Memo. 467. http://www.ecmwf.int/publications/library/do/references/show?id=86713

Sun, Juanzhen and N. Andrew Crook, 1997: "Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part I: Model Development and Simulated Data Experiments" *J. Atmos. Sci*., **54**, 1642-1661

Sun, Juanzhen, 2006: "Convective-scale Data Assimilation: Progress and Challenges" *Quart. J. Roy. Met. Soc. Special Issue devoted to papers from the 4th WMO Symposium on Data Assimilation.*

Tanguay, Monique, Peter Bartello and Pierre Gauthier, 1995: "Four-dimensional data assimilation with a wide range of scales", *Tellus*, **47A**, 974-997.

Tribbia, J. J., Baumhefner, D. P. 2004: "Scale Interactions and Atmospheric Predictability: An Updated Perspective" *Mon. Wea. Rev.* **132**, 703-713.

WCRP 2005: "The WCRP Strategic Framework 2005-2015 Coordinated Observation and Prediction of the Earth System (COPES)" WCRP-123 WMO/TD-No291 http://copes.ipsl.jussieu.fr/PDF/WCRPStrategImpl/WCRP_StrateIgmpl.PDF

WGNE 1996: Report of 11th session of the CAS/JSC Working Group on Numerical Experimentation, Nov 1995. *CAS/JSC WGNE Report No 11.*

Xu, Qin 1996: "Generalized Adjoint for Physical Processes with Parameterized Discontinuities. Part I: Basic Issues and Heuristic Examples" *J. Atmos. Sci.*, **53**, 1123-1142

# Figures

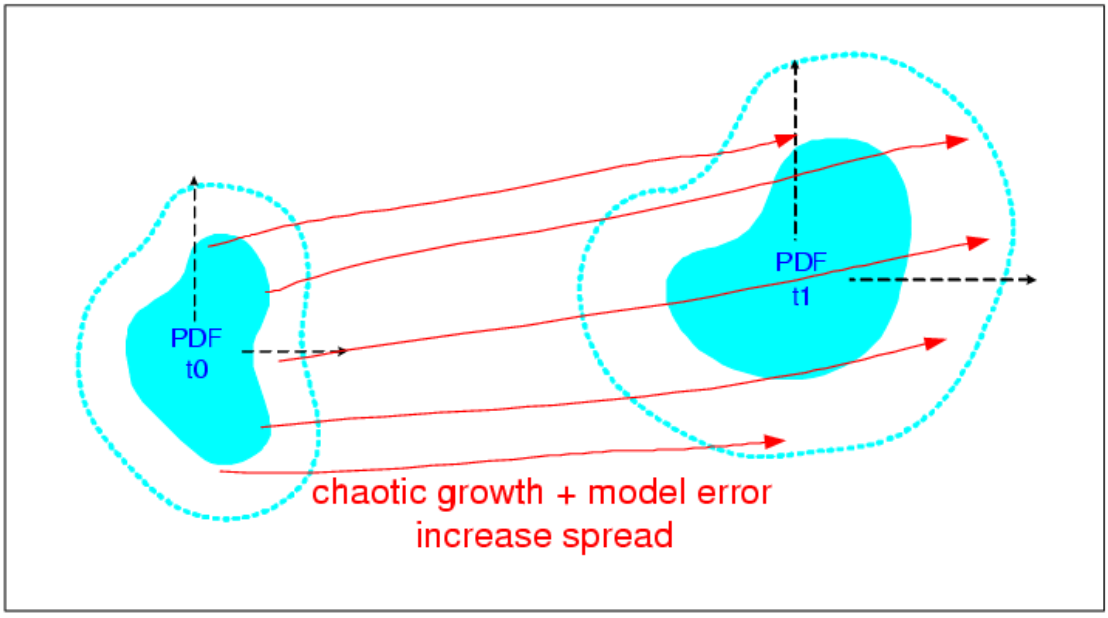


**Figure 1 Sketch of the ideal evolution of a PDF using the Fokker-Planck equation.**

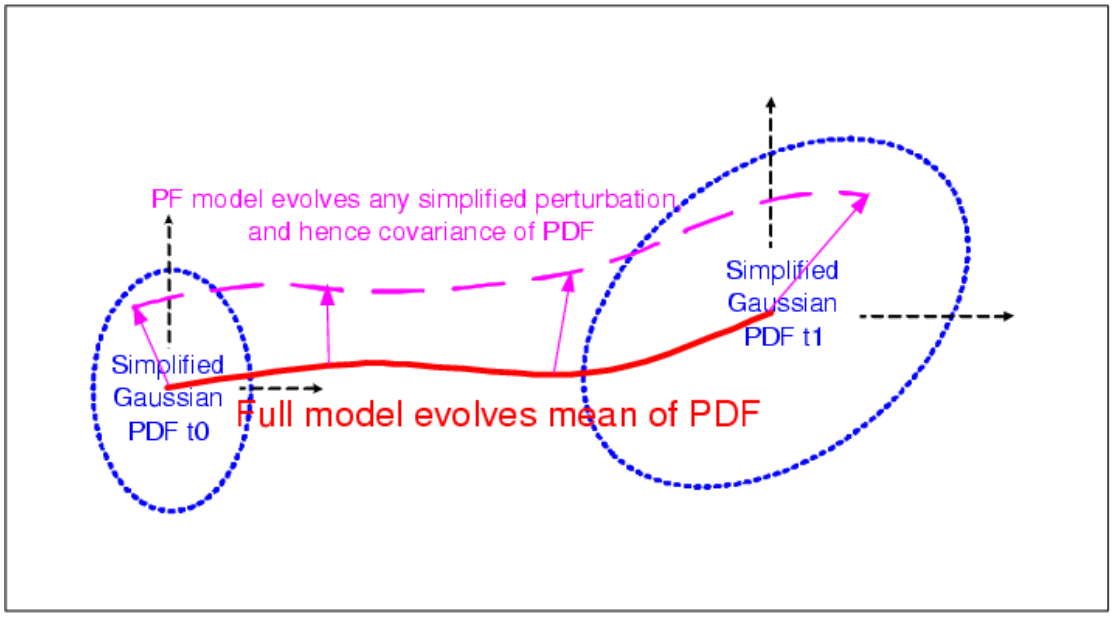


**Figure 2 Approximation to figure 1, using Gaussians, and evolving the mean and linear perturbations.**